Classifying Faces with Non-negative Matrix Factorization

David Guillamet , Jordi Vitrià

Computer Vision Center, Dept. Informàtica
Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, Spain
{davidg,jordi}@cvc.uab.es

Abstract

This paper addresses the well-known problem of recognizing faces under several unfavorable situations. We have analyzed situations with changes in expression, in illumination and occlusions such as faces wearing sunglasses or scarfs. We have introduced the use of the Non-negative Matrix Factorization (NMF) technique in the context of classification of face images and we have directly compared performances of NMF and Principal Component Analysis (PCA) using a well-known face database, the AR, that contains a large number of individuals taken under several conditions. Moreover, these results have also been compared to two leading algorithms, one template based and the other feature based, noticing that NMF is able to improve them when using a high dimensional space. In addition, NMF has been used with some distance metrics as L1, L2 or correlation in order to determine the best one for such problem. We have discovered that the correlation metric is the most suitable one for our problem.

Keywords: Computer vision.

1 Introduction

Face recognition is one of the most challenging problems to be solved in the computer vision community. Until now, several methods and sophisticated approaches have been developed in order to obtain the best recognition results using some specific face databases. Due to this huge number of methods and face databases, there is no uniform way to establish the best method just because nearly all of them have been designed to work with some specific face situations. Even though, some of these methodologies have lead to the development of a great number of commercial face recognition systems. Most of the face recognition algorithms can be classified into two classes, image template based or geometry feature based. Template based methods compute a measure of correlation between new faces and a set of template models to estimate the face identity. Several well-known statistical techniques have been used to define a template model, such as Support Vector Machines (SVM) [12], Linear Discriminant Analysis (LDA) [1], Principal Component Analysis (PCA) [11] and Independent Component Analysis (ICA) [2]. Usually, these approaches are focused on extracting global face features, and occlusions are difficult to handle. Geometry feature-based methods analyze explicit local facial features, and their geometric relationships. Some examples of these methods are the active shape model [4], the elastic bunch graph matching algorithm for face recognition [13] and the Local Feature Analysis (LFA) [10].

In this paper we address the problem of recognizing frontal faces captured in different illumination conditions and containing natural occlusions such as individuals wearing sunglasses and scarfs. We have to note that the problem of recognizing faces under natural occlusions is a must if we are developing a robust face classifier. In order to obtain comparable results with the most important techniques, we have used a face database that has been extensively used by the computer vision community, the AR face database [7]. Furthermore, in this paper we introduce the Non-negative Matrix Factorization (NMF) [5, 6] technique in a face classification framework noticing its ability to deal with natural occlusions. As NMF is based on a subspace definition, we have also introduced the Principal Component Analysis (PCA) for a direct comparison. We also present present some preliminary results concerning to the determination of which distance metric should be used in the feature space created by the positive restrictions of NMF. In order to evaluate the introduction of NMF in such a framework, we have taken as a reference the results of a previous work [3] that used the same face database for analyzing the two leading commercial face recognition techniques, Local Feature Analysis and Bayesian PCA.

2 PCA and NMF techniques

2.1 Principal Component Analysis

Due to the high dimensionality of data, similarity and distance metrics are computationally expensive and some compaction of the original data is needed. Principal Component Analysis is an optimal linear dimensionality reduction scheme with respect to the mean squared error (MSE) of the reconstruction. For a set of N training vectors $X = \{x^1, \ldots x^N\}$ the mean ($\mu = \frac{1}{N}\sum_{i=1}^{N} x^i$) and covariance matrix ($\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x^i\mu)(x^i - \mu)^T$) can be calculated. Defining a projection matrix E composed of the K eigenvectors of Σ with highest eigenvalues, the K-dimensional representation of an original, n-dimensional vector x, is given by the projection $y = E^T(x - \mu)$.

2.2 Non-Negative Matrix Factorization

NMF is a method to obtain a representation of data using non-negativity constraints. These constraints lead to a part-based representation because they allow only additive, not subtractive, combinations of the original data [5]. Given an initial database expressed by a $n \times m$ matrix V, where each column is an n-dimensional non-negative vector of the original database (m vectors), it is possible to find two new matrices (W and H) in order to approximate the original matrix $V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia}H_{a\mu}$. The dimensions of the factorized matrices W and H are $n \times r$ and $r \times m$, respectively. Usually, r is chosen so that $(n + m)r < nm$. Each column of matrix W contains a basis vector while each column of H contains the weights needed to approximate the corresponding column in V using the bases from W. In the PCA context, each column of matrix W represents an eigenvector and the factorized matrix of H represent the eigenprojections. In contrast to PCA, NMF does not allow negative entries in the factorized matrices W and H permitting the combination of multiple bases images to represent an object.

In order to estimate the factorization matrices, an objective function has to be defined. A possible objective function is given by $F = \sum_{i=1}^{n}\sum_{\mu=1}^{m}[V_{i\mu}\log(WH)_{i\mu} - (WH)_{i\mu}]$. This objective function can be related to the likelihood of generating the images in V from the bases W and encodings H. An iterative approach to reach a local maximum of this objective function is given by the following rules [5]: $W_{ia} \leftarrow W_{ia}\sum_{\mu}\frac{V_{i\mu}}{(WH)_{i\mu}}H_{a\mu}$, $W_{ia} \leftarrow \frac{W_{ia}}{\sum_j W_{ja}}$, $H_{a\mu} \leftarrow H_{a\mu}\sum_i W_{ia}\frac{V_{i\mu}}{(WH)_{i\mu}}$. Initialization is performed using positive random initial conditions for matrices W and H. The convergence of the process is also ensured. See [5, 6] for more information.

3 Experimental results

Our experiments are based on the direct comparison of principal component analysis (PCA) and non-negative matrix facortization (NMF) techniques in a well-known face database. Furthermore, the obtained results are compared to two leading techniques used in the computer vision community: the FaceIt and Bayesian techniques. FaceIt technique is a successful commercial face recognition system and it is mainly based on the Local Feature Analysis (LFA) [10]. The Bayesian technique was developed by Moghaddam and Pentland [9] in order to model large non-linear variations in facial apperance due to self-occlusion and self-shading using a PCA approach as a probability density estimation tool. See [3] for more detailed information about the comparison of these two techniques in the AR face database. Gross et al. [3] have compared both techniques using the AR face database by considering a wide set of situations: natural occlusions, changes in expression, changes in the lighting conditions.

Our current work compares the performance obtained by the NMF with PCA and these two techniques. Furthermore, we have also analyzed different distance metrics in the NMF projected space to take into account how a suitable metric can affect to the classification results.

3.1 AR database

The AR database was collected at the Computer Vision Center in Barcelona [7] and it contains images of 116 individuals (63 males and 53 females). Original images are 768×576 pixels in size with 24-bit color resolution. This database is very interesting because subjects were recorded twice at a 2-week interval. And during each session 13 conditions with varying facial expressions, illumination and occlusion were captured. This means that each individual will contain some local variations in appearance. For example, some individuals are captured wearing normal glasses during the first session but not in the next.

Due to high dimensionality of original images, we reduced them to a 40×48 size where our representation becomes more manageable. As our two statistical techniques will be based on a pixel representation of each image, a pose normalization has been applied in order to align all database faces. We have manually localized both eye positions in every image and we have normalized all faces according to this information. Moreover, in order to avoid external influences of background, we have only considered the part of a face inside an elliptical region. Figure (1) shows an example of an individual taken under different conditions and the elliptical region considered. The size of each reduced image is 40×48 and if we consider the elliptical region, each image is represented using 1505 pixels. The elliptical region considered has been extracted after analyzing all images in the database and rejecting all those pixels that do not have a statistical influence in the face.

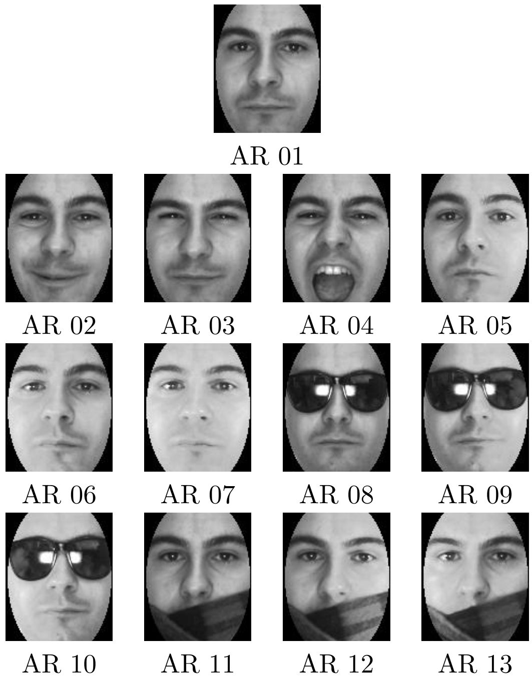


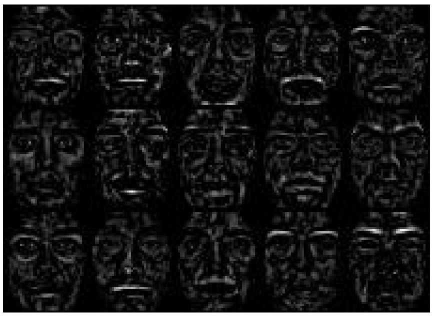
Figure 1: AR database. Example of one individual of the AR face database that reflects all the possible situations. The conditions are: (1) neutral, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both lights on, (8) sunglasses, (9) sunglasses/left light, (10) sunglasses/right light, (11) scarf, (12) scarf/left light, (13) scarf/right light. Each image is of size 40×48 and the size considered for our experiments is 1505 when we consider the elliptical region.

3.2 Evaluation of algorithms

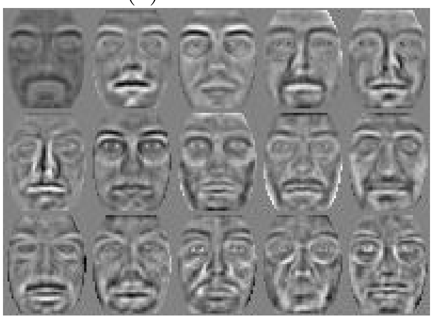
Faces are projected in a low dimensional space that in this particular study is limited to 50, 100 and 150 dimensions in order to have a general idea of how results can change when the number of dimensions of the feature space is modified. As known, Non-negative Matrix Factorization is a part based technique and Principal Component Analysis a global one and this behaviour is reflected in the bases obtained by both techniques. Figure (2) shows some of the bases where we can initially see that NMF provides a more sparse representation instead of the global one provided by PCA.

Each individual situation is analyzed when both PCA and NMF algorithms are used for classification. Thus, we can compare when each technique is performing better and try to understand when can be used for a further classification task. The training images consist of two neutral poses of each individual that were captured in two different days. In figure (1) we can see a training sample that is labelled as AR 01.

In order to see how each technique can deal with expressions, images labelled as AR 02, 03 and 04 are used as a testing set because they contain smile, anger and scream expressions. Table (1) shows the results of both algorithms with respect to the FaceIt and Bayesian techniques. The first impression is that L2 Norm is not the most suitable metric when working with Non-negative Matrix Factor-



(a) NMF bases.



(b) PCA bases.

Figure 2: Bases obtained by both techniques, PCA and NMF. NMF provides a more sparse and part based representation and PCA a global one.

ization and both L1 Norm and correlation metrics could be a good choice. Expression AR 02 is better classified by FaceIt and AR 03 is better classified when using NMF in a high dimensional space. But expression AR 04 demonstrates that is a very difficult expression where PCA and NMF are not able to deal with.

		Expression		
		AR 02	AR 03	AR 04
FaceIt		0.96	0.93	0.78
Bayesian		0.72	0.67	0.41
PCA-50+L2 Norm		0.67	0.82	0.18
NMF-50 +	L2 Norm	0.61	0.78	0.14
	L1 Norm	0.72	0.80	0.19
	Correlation	0.73	0.77	0.18
PCA-100+L2 Norm		0.80	0.88	0.24
NMF-100 +	L2 Norm	0.62	0.85	0.09
	L1 Norm	0.85	0.91	0.29
	Correlation	0.89	0.90	0.28
PCA-150+L2 Norm		0.83	0.90	0.29
NMF-150 +	L2 Norm	0.66	0.87	0.09
	L1 Norm	0.88	0.92	0.30
	Correlation	0.93	0.95	0.36

Table 1: Expression results. This table reflects how both techniques can deal with facial expressions. Note that scream expression (AR 04) is hard to recognize.

Illumination conditions are also a factor to take into account in a face recognition framework. This conditions are reflected in images AR 05, AR 06 and AR 07. Table (2) shows that PCA can not deal with illumination conditions as good as the NMF and when the number of dimensions starts to increase, NMF can improve the FaceIt and Bayesian approaches.

		Expression		
		AR 05	AR 06	AR 07
FaceIt		0.95	0.93	0.86
Bayesian		0.77	0.74	0.72
PCA-50+L2 Norm		0.77	0.76	0.57
NMF-50 +	L2 Norm	0.91	0.84	0.67
	L1 Norm	0.93	0.87	0.69
	Correlation	0.94	0.89	0.76
PCA-100+L2 Norm		0.86	0.86	0.69
NMF-100 +	L2 Norm	0.94	0.85	0.67
	L1 Norm	0.97	0.94	0.87
	Correlation	0.99	0.94	0.88
PCA-150+L2 Norm		0.85	0.87	0.71
NMF-150 +	L2 Norm	0.93	0.84	0.64
	L1 Norm	0.98	0.97	0.92
	Correlation	0.99	0.96	0.91

Table 2: Illumination results. This table reflects how both techniques can deal with illumination changes. Note that in this case, NMF obtains the best classification results when the number of dimensions starts to be considerable (100 or 150).

Occlusions have been considered a topic of research in the computer vision community. Here, we have a set of natural occlusion where faces are occluded with a scarf and sunglasses. AR 08 contains sunglasses that occlude both eyes and AR 09 and AR 10 consider the same situation but including left and right illuminations. So, it will be expected that recognition rates for these situations start to decrease. AR 11 images consider a scarf and that means that mouth is occluded. AR 12 and AR 13 also consider a scarf but with the addition of illumination conditions. Tables (3) and (4) show all the results obtained when considering these two kinds of occlusions.

Under the presence of sunglasses, recognition rates decrease considerably as can be seen in table (3). This means that eyes are a very important feature to take into consideration when classifying faces. It is interesting to note that when sunglasses are considered without considering lighting influences (AR 08), NMF obtains the best recognition results. But, when lighting conditions are present, the Bayesian technique gives the best results. Thus, NMF is a good choice when partial occlusions are present but when lighting conditions affect to the

		Expression		
		AR 08	AR 09	AR 10
FaceIt		0.10	0.08	0.06
Bayesian		0.34	0.35	0.28
PCA-50+L2 Norm		0.16	0.12	0.18
NMF-50 +	L2 Norm	0.16	0.10	0.12
	L1 Norm	0.19	0.10	0.20
	Correlation	0.23	0.12	0.17
PCA-100+L2 Norm		0.23	0.15	0.22
NMF-100 +	L2 Norm	0.14	0.11	0.12
	L1 Norm	0.24	0.15	0.21
	Correlation	0.32	0.19	0.24
PCA-150+L2 Norm		0.26	0.16	0.24
NMF-150 +	L2 Norm	0.17	0.12	0.09
	L1 Norm	0.31	0.21	0.23
	Correlation	0.38	0.21	0.23

Table 3: Occlusion results when considering sunglasses. Note that in this case, NMF only is better when using a high dimensional feature space and no lighting conditions are considered. When lighting conditions are considered, Bayesian approach obtains the best recognition rates.

		Expression		
		AR 11	AR 12	AR 13
FaceIt		0.81	0.73	0.71
Bayesian		0.46	0.43	0.40
PCA-50+L2 Norm		0.44	0.38	0.37
NMF-50 +	L2 Norm	0.47	0.35	0.28
	L1 Norm	0.59	0.35	0.32
	Correlation	0.61	0.45	0.35
PCA-100+L2 Norm		0.59	0.50	0.47
NMF-100 +	L2 Norm	0.47	0.36	0.25
	L1 Norm	0.66	0.55	0.46
	Correlation	0.76	0.62	0.59
PCA-150+L2 Norm		0.62	0.57	0.48
NMF-150 +	L2 Norm	0.53	0.31	0.24
	L1 Norm	0.73	0.57	0.48
	Correlation	0.75	0.62	0.56

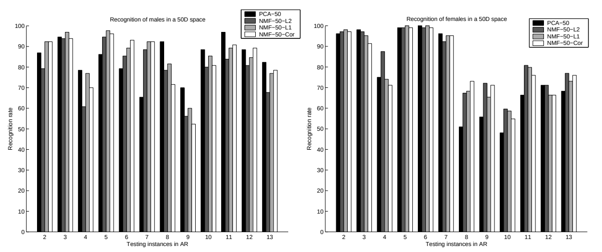
Table 4: Occlusion results when considering a scarf. In this case, FaceIt is obtaining the best recognition results. And we have to note that NMF is better than the Bayesian approach in this situation. Again, NMF is always better than PCA.

scene, it turns out that NMF can not deal with a more general change in the scene. Table (4) shows a similar behaviour with the NMF, when no lighting conditions are present (AR 11) in the scene, NMF can have a high recognition rate, even comparable to the best one obtained with FaceIt, but when lighting conditions are present (AR 12 and AR 13) all recognition rates decrease considerably.

In general, the first impression of these first experiments is that NMF performs better than PCA in the same dimensional space. This behaviour was expected because PCA is based on a global transformation of the original space and NMF on a local one. NMF is based on representing the original space using a set of bases that, as demonstrated by Lee and Seung [5], are parts of objects. Thus, it turns out that when we are considering local effects as occlusions, changes in expression or even changes in the illumination, PCA is not able to represent them as well as NMF. In terms of performances with respect to the FaceIt and Bayesian techniques, NMF has comparable recognition rates and, in some situations, is even better than these two methods. The reason of this high performance is mainly justified by its natural definition of representing data using a combination of bases that are part-based. Finally, it is clear that L2 is the worst metric to use with NMF and the correlation metric is the best one.

3.3 Gender classification

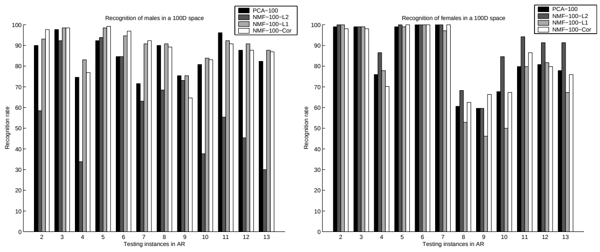
It is clear that if we try to distinguish between males and females, local features corresponding to each gender are different. Thus, this means that Non-negative Matrix Factorization (NMF) can be a suitable technique for capturing these local differences. This motivates to create a gender classifier based on the NMF and when a testing face is correctly classified according to its gender, we can use this information to recognize the face using a more specific face classifier. In our study, we have learned two gender classifiers: one with PCA and the other with NMF with the same parameters as in the previous experiments. Figures (3), (4) and (5) show the gender classification results when using a 50 dimensional space (figure (3)), a 100 dimensional space (figure (4)) and a 150 dimensional space (figure (5)).



(a) (b)

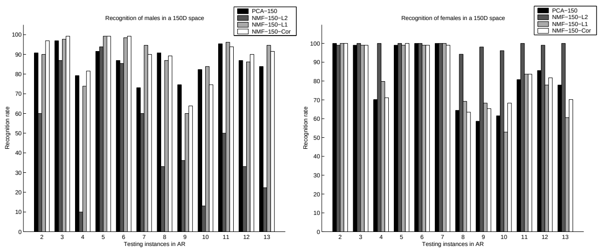
Figure 3: Gender classification results when trying to classify males (a) and females (b) using a 50 dimensional feature space.

Figures (3), (4) and (5) depict a general behaviour for both PCA and NMF techniques: females are better recognized in this set of situations: AR02,AR03,AR05,AR06,AR07 and males in



(a) (b)

Figure 4: Gender classification results when trying to classify males (a) and females (b) using a 100 dimensional feature space.



(a) (b)

Figure 5: Gender classification results when trying to classify males (a) and females (b) using a 150 dimensional feature space.

the other ones . These recognition differences must be studied more deeply but this means that both genders have some local features that are better identified depending on the face situation.

Images with occlusions, as AR 8, AR 9 and AR 10 are very difficult to identify even when trying to determine whether it is a male or female. But, in general, recognition rates are very high. Thus, we can take this gender information to train a face classifier only based on males or females. We have to consider that NMF is based on capturing local behaviours, so, a more specific classifier based only on males or females should improve the initial recognition rates presented before. Once a gender classifier is trained with the same training instances as for the previous experiments for each PCA and NMF techniques, we have repeated all experiments in order to determine whether a gender classifier is convenient or not. These experiments are reflected in tables 5,6,7).

In general, with the addition of a gender classifier both techniques (PCA and NMF) are slightly improved. This improvement is not very significant in face images containing hard occlusions such as

		Expression		
		AR 02	AR 03	AR 04
PCA-50+L2 Norm		0.74	0.87	0.22
NMF-50 +	L2 Norm	0.68	0.83	0.17
	L1 Norm	0.81	0.87	0.25
	Correlation	0.85	0.84	0.25
PCA-100+L2 Norm		0.83	0.91	0.28
NMF-100 +	L2 Norm	0.65	0.87	0.14
	L1 Norm	0.90	0.92	0.31
	Correlation	0.91	0.94	0.34
PCA-150+L2 Norm		0.84	0.91	0.28
NMF-150 +	L2 Norm	0.70	0.88	0.13
	L1 Norm	0.90	0.93	0.33
	Correlation	0.93	0.94	0.35

Table 5: Expression results when considering a previous gender classifier. This table must be compared with table (1) where we can appreciate some improvements.

		Expression		
		AR 05	AR 06	AR 07
PCA-50+L2 Norm		0.82	0.82	0.62
NMF-50 +	L2 Norm	0.91	0.86	0.68
	L1 Norm	0.94	0.89	0.76
	Correlation	0.96	0.92	0.84
PCA-100+L2 Norm		0.86	0.86	0.70
NMF-100 +	L2 Norm	0.92	0.89	0.73
	L1 Norm	0.97	0.96	0.89
	Correlation	0.98	0.97	0.92
PCA-150+L2 Norm		0.86	0.87	0.71
NMF-150 +	L2 Norm	0.94	0.89	0.69
	L1 Norm	0.98	0.98	0.92
	Correlation	0.98	0.97	0.93

Table 6: Illumination results when considering a previous gender classifier. This table must be compared with table (2). In this particular case, AR 07 is specially improved in low dimensional spaces.

		Expression		
		AR 08	AR 09	AR 10
PCA-50+L2 Norm		0.21	0.13	0.20
NMF-50 +	L2 Norm	0.20	0.10	0.14
	L1 Norm	0.24	0.14	0.22
	Correlation	0.29	0.17	0.24
PCA-100+L2 Norm		0.25	0.16	0.23
NMF-100 +	L2 Norm	0.16	0.15	0.13
	L1 Norm	0.26	0.17	0.21
	Correlation	0.35	0.21	0.25
PCA-150+L2 Norm		0.27	0.17	0.25
NMF-150 +	L2 Norm	0.18	0.13	0.10
	L1 Norm	0.32	0.20	0.24
	Correlation	0.36	0.24	0.26

Table 7: Occlusion results when considering sunglasses and a previous gender classifier. This table must be compared with table (3). We can see that in this particular case, recognition rates are not really improved.

those faces containing sunglasses or a scarf. However, these results motivate to build up a face classifier divided into a global gender detector and two specific face classifiers, one for males and another

		Expression		
		AR 11	AR 12	AR 13
PCA-50+L2 Norm		0.52	0.44	0.40
NMF-50 +	L2 Norm	0.51	0.34	0.31
	L1 Norm	0.60	0.41	0.37
	Correlation	0.67	0.51	0.45
PCA-100+L2 Norm		0.64	0.55	0.49
NMF-100 +	L2 Norm	0.50	0.32	0.24
	L1 Norm	0.71	0.56	0.51
	Correlation	0.79	0.62	0.57
PCA-150+L2 Norm		0.63	0.57	0.51
NMF-150 +	L2 Norm	0.53	0.36	0.27
	L1 Norm	0.73	0.58	0.52
	Correlation	0.79	0.65	0.61

Table 8: Occlusion results when considering a scarf and a previous gender classifier. This table must be compared with table (4). In this case, there is a general improvement in recognition rates.

for females. This configuration must work out much more better than only considering a face classifier because NMF is based on the representation of local features. Figure (6) summarizes previous results showing all the recognition rates obtained according to the method used (PCA or NMF) in conjunction with their internal parameters. We have to note that the overall recognition rate for the FaceIt technique is 65.83% and 52.42% for the Bayesian one.

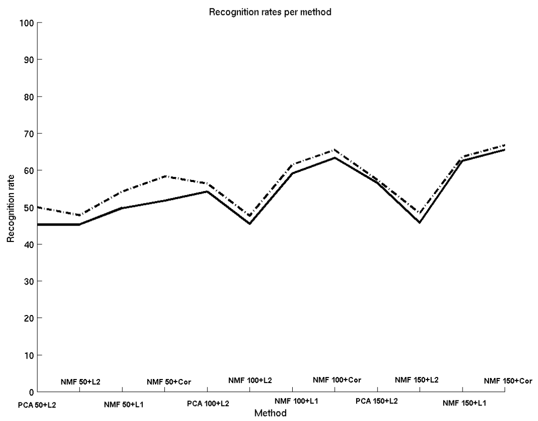


Figure 6: Recognition rates according to the method used. Solid line indicates the recognition rates obtained without using any gender information and the dashed line indicates the recognition rates when considering gender information. The best method according to the whole set of face situations is the Non-negative Matrix Factorization in a 150 dimensional space using the correlation distance as a metric obtaining a recognition rate of 66.74%.

From the analysis of figure (6), we can appreciate that the introduction of a gender classifier improves the whole recognition rates even using PCA or NMF. Obviously, this behaviour is justified because it is more easy to classify a face into a male or female than recognizing the face directly. But it is clear that this improvement is more remarkable in low dimensional spaces.

If we directly compare the overall results obtained using PCA and NMF with respect to the FaceIt and Bayesian techniques, we can state that performances are comparable depending on the low dimensional space. The best configuration of our scheme is the one that uses the Non-negative Matrix Factorization in a 150 dimensional space using the correlation metric where we obtain a recognition rate of 66.74 that is greater than the recognition rate of 65.83% obtained by the FaceIt technique. Of course, this behaviour is not observed in all the face situations but it is clear that for some certain situations one technique will be more convenient than another one.

4 Conclusions

In this paper we have analyzed the Non-negative Matrix Factorization (NMF) technique in the problem of recognizing faces captured under non favorable conditions such as changes in expressions, changes in the lighting conditions and occlusions. Results are also compared with the well-known Principal Component Analysis (PCA) technique because both algorithms are based on finding a subspace where our data can be expressed. In the particular case of NMF, this subspace is defined by positive restrictions and, as noted by Lee and Seung [5], it is able to find part based decompositions of the data. Our experiments have demonstrated that this specific feature of NMF allows for high recognition rates in comparison with PCA, which treats its input data in a global a way. And it is clear that these results are justified by the fact that our face database contains local variations of faces, not global ones. As NMF is a recent technique, no distance metrics have been defined for its positive subspace. In our study, we have analysed L1, L2 and correlation distances noticing that the last is the most suitable one. A gender classifier preclassification stage has also been introduced in order to obtain the best results.

Finally, we have compared our results to the two leading face recognition techniques (FaceIt and Bayesian), noticing that our scheme is more adapted to the problem of recognizing faces under several unfavorable conditions. This is justified by the fact that these two techniques have been designed to work with faces that contain specific changes in expressions, but not the whole range of conditions that we have exposed in this current work.

References

[1] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711-720, July, 1997.

[2] P. Comon. Independent component analysis - a new concept ? *Signal Processing* 36:287-314, 1994.

[3] R. Gross, J. Shi and J. F. Cohn. Quo vadis Face Recognition? In Proc. third Workshop on Empirical Evaluation Methods in Computer Vision, December, 2001.

[4] A. Lanitis, C. J. Taylor and T. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743-756, 1997.

[5] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788-791, 1999.

[6] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. *NIPS*, 2000.

[7] A. Martínez and R. Benavente. The AR face database. *Technical report 24, Computer Vision Center (CVC)*. Barcelona, Spain, June 1998.

[8] B. Moghaddam. Principal Manifolds and Probabilistic Subspaces for Visual Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(6), June 2002.

[9] B. Moghaddam and A.P. Pentland. Probabilisitc visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696-710,1997.

[10] P. Penev and J. Atick. Local feature analysis: A general statistical theory for object representation, 1996.

[11] M. Turk and A.P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991.

[12] V. N. Vapnik. *The nature of statistical learning theory.* Springer Verlag, Heidelberg, DE, 1995.

[13] L. Wiskott, J.M. Fellous, N. Kruger and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(7):775-779, July 1997.